On the Benefits of Flexible Customer-to-Depot Assignments in Attended Home Delivery

WORKING PAPER
April 2019

Kilian Seifried
Logistics and Supply Chain Management, Business School, University of Mannheim, Germany
kilian.seifried@bwl.uni-mannheim.de

Niels Agatz
Technology and Operations Management, Rotterdam School of Management, Erasmus University, The Netherlands
nagatz@rsm.nl

Moritz Fleischmann
Logistics and Supply Chain Management, Business School, University of Mannheim, Germany
mfleischmann@bwl.uni-mannheim.de

Abstract
This paper investigates the value of flexibility in customer-to-depot assignments in last-mile distribution operations. In current practice, many companies use fixed delivery regions for their fulfillment warehouses. This allows a decomposition of delivery operations by warehouse, thereby simplifying the planning process. However, this approach foregoes some optimization potential, relative to a holistic planning of the delivery operations across the network.

In this paper, we assess the value of this flexibility. We first identify and characterize two specific types of benefits, namely a routing benefit and a pooling benefit. We then proceed to quantify these benefits in an extensive numerical study. Our results disentangle how the benefits depend on key environmental parameters. They also show that a flexible customer-to-depot assignment can add significant value in various relevant conditions.

Keywords
Routing – multiple depots – time windows – attended home delivery
1 Introduction

This paper investigates the value of flexibility in customer-to-depot assignments in last-mile distribution operations. The insights obtained are relevant to e-commerce businesses using attended home delivery (AHD), namely e-groceries.

Global e-grocery sales have substantially increased in recent years, and this trend is likely to continue. For example, the online grocery market in the UK is expected to grow from 15 billion USD in 2015 to 28 billion in 2020 (Cahill, 2016), while in the US the market is expected to grow to 18 billion USD (Cahill, 2016). The largest e-grocery market is China, with a predicted volume of 178 billion USD in 2020 (Cahill, 2016).

The online grocery landscape consists of pure online players such as Fresh Direct (US), Picnic (NL/DE), Ocado (UK), and Amazon Fresh (multinational) and traditional bricks-and-mortar retailers such as TESCO (UK), Albert Heijn (NL) and REWE (DE). In both cases, AHD is the dominant service model. This model requires the e-grocer and the customer to agree on a time window for delivery. The menu of delivery time windows offered is a major part of the customer’s service experience. At the same time, it has a strong impact on the e-grocer’s delivery efficiency. The design of the time window menu has to address the trade-off between these effects in sales and in operations (cf. Agatz et al., 2013).

To allow for efficient order picking, orders are usually picked and packed in dedicated e-fulfillment warehouses. Efficient picking is important for groceries as basket sizes are large and typically consist of many different stock keeping units, including items that require special care, such as cooling. Then, the orders are delivered from the e-fulfillment warehouse to the customers, either by a third party provider or the e-grocer’s own vehicles. Many online grocers deploy their own fleet as they consider it crucial to keep the delivery under their control as a central customer touch point (Weikert, 2017).

To limit transportation costs and allow for short delivery lead times, most e-grocers use multiple fulfillment warehouses. Each warehouse typically covers a fixed geographic delivery region. Dutch grocery retailer Albert Heijn, for example, serves most of the Netherlands from five e-fulfillment centers that each cover a specific delivery area (Silicon Canals, 2019). Another example is pure online grocer Ocado that operates four customer distribution centers and a larger number of hubs in the UK to ensure same and next-day deliveries to its British customers (Ocado Group plc, 2017).

In summary, effective and cost-efficient home-delivery of groceries is highly challenging, due to thin margins, high customer expectations, and extensive last-mile operations. Many e-grocers struggle to attain profitability, and there is no established road to success. Therefore, e-grocers are continuously looking for ways to improve the efficiency of their last-mile delivery operations and to better balance their cost-service performance.

In this paper, we investigate how an e-grocer’s network of fulfillment warehouses can serve as a source of flexibility. Specifically, we investigate the benefits of a flexible assignment of customer orders to fulfillment warehouses. The current practice of using fixed delivery regions allows a decomposition of delivery operations by warehouse. This simplifies the planning process. On the
other hand, this approach foregoes some optimization potential, relative to a holistic planning of the delivery operations across the entire network. In this study, we first characterize the specific benefits of such a holistic optimization of delivery operations in a fulfillment network. We then proceed to quantify these benefits and show how they depend on key environmental parameters.

A flexible assignment of customer orders to fulfillment locations is common for some online retailers outside of the grocery business. Several studies in the literature investigate the resulting inventory pooling benefits (Xu et al., 2009; Acimovic and Graves, 2015; Torabi et al., 2015). Similar effects may arise in e-groceries. What is different, however, is the last-mile delivery process. While in the above examples shipping is, in general, outsourced to a logistics service provider, e-grocers typically use their own fleet to bridge the last-mile. This adds routing considerations to the delivery planning. We investigate how this routing impact affects the value of a flexible order-to-warehouse assignment.

Conceptually, determining optimal delivery routes under a flexible order assignment leads to a multi-depot vehicle routing problem with time windows (MDVRPTW) (Cordeau et al., 2001). As exact approaches can only solve small instances (Dondo et al., 2003; Dondo and Cerdá, 2007; Bae and Moon, 2016), most work in this area focuses on heuristics, especially tabu search (Cordeau et al., 2001, 2004) and variable neighborhood search (VNS) (Polacek et al., 2004, 2008). In this paper, we apply a version of this VNS heuristic in combination with a new mixed integer programming model in our experiments. For a comprehensive overview of the literature on the MDVRPTW, we refer to the survey of Montoya-Torres et al. (2015) and the more general survey on 'rich' vehicle routing by Caceres-Cruz et al. (2015).

While a lot of research is dedicated to solution approaches for multi-depot routing problems, we are not aware of research that specifically investigates the benefits of an integrated planning approach as compared to decomposing the problem by depot. With this paper, we aim to address this gap and analyze the benefits of flexible customer-to-depot assignments, especially for the operation of AHD in the e-grocery business. In particular, this paper makes the following contributions: (i) We identify different types of benefits of a flexible customer-to-depot assignment; (ii) we quantify these benefits in different environments by conducting a wide range of numerical experiments; (iii) we present a novel mixed integer programming (MIP) formulation that requires much fewer variables than previously available formulations.

The remainder of the paper is organized as follows. We motivate the development of our model and our experiments in Section 2. In Section 3, we formulate the problem as a two-index MIP model. We then investigate the benefits of the flexible assignments in Section 4. Finally, in Section 5, we summarize our most important insights and give a short outlook.

2 The Attended-Home-Delivery Planning Process

To understand the effects of flexibility in the order-to-depot assignment, it is helpful to look at the planning process of attended home delivery operations. In this section, we therefore summarize this process. Figure 1 depicts the main planning phases, following Hungerländer et al. (2018) and Agatz et al. (2008). We briefly discuss each of these phases in the context of
e-grocery. We start at the tactical planning level since we take strategic decisions, such as the choice of warehouse locations, as input to our analysis.

On the tactical level, the e-grocer determines, for each depot, the size and composition of the fleet and the associated workforce. Furthermore, she defines a menu of possible delivery time windows. In current practice, the depot assignment is also determined at this tactical level in the sense that each depot typically serves a fixed delivery region. As a consequence, the subsequent operational planning steps can be decomposed by depot.

The ordering phase starts several weeks before delivery, with customers placing orders online. For each incoming customer order, the e-grocer may dynamically update the offered time windows, based on the available capacity (in picking and in routing). This can be feasibility-oriented, based on already accepted orders (can we still serve this order?) or profit-oriented, taking into account projected future orders (do we want to serve this order or reserve capacity for future customers?). For each delivery shift, orders are accepted until a cut-off time, e.g., midnight for next-day delivery.

In the preparation phase, after the cut-off time, the e-grocer determines the routes to deliver all accepted orders. Subsequently, the orders are picked and packed in the depots. The planning problems of the ordering phase and the preparation phase are interlinked as the available capacity in the ordering phase depends on the routing decisions in the preparation phase.

After the preparation phase, the retailer executes the planned delivery routes to serve all customers in their selected time windows.

The flexibility that we assess in this paper stems from moving the customer-to-depot assignment from the tactical planning phase to the preparation phase, i.e., after the order cut-off point. The assignment can then be made based on actual orders, rather than projected orders. In that case, the depot assignment is made concurrently with the route planning. This gives rise to additional complexities in operational planning and execution. It is, for example, more difficult to solve the larger vehicle routing problems associated with planning across depots. At the same time, a flexible depot assignment yields several potential benefits. We highlight the
two following benefits and focus on them in our analysis.

- In the preparation phase, the objective is to serve all accepted customers at minimum cost. The multi-depot version of the vehicle routing problem relaxes the constraints that each customer should be served on a route that originates from a pre-defined depot location. This yields additional routing flexibility that can help to reduce routing costs. We refer to this type of benefit as *routing benefit*.

- In the ordering phase, the objective is to serve as many customers as possible, given the available routing and picking capacity. Consolidating the capacities of the different depot locations may lead to *pooling benefits*, i.e., allowing a higher overall capacity utilization.

Both of these effects interact since the order acceptance decision defines the requirements for the route planning. For an economically viable AHD operation, the overall goal should be to maximize the operational profit, i.e., the revenue generated by accepting customer orders minus the costs for executing the delivery tours. Thus, the decisions of order acceptance and routing should be integrated, and the depot assignment plays a role in both of them. The benefit of a flexible depot assignment is then the gain in the operational profit, combining saved routing costs and increased revenues from more customers served. We refer to this metric as the *combined benefit* of a flexible depot assignment.

In Section 4, we evaluate each of these types of benefits. As a basis for our analysis we develop, in Section 3, a MIP model for the profit-maximizing MDVRPTW that captures the integration of order acceptance and routing decisions and allows us to analyze the combined benefit. We also present two model variants that allow us to isolate the routing and pooling benefits, respectively.

### 3 Model Formulation

In this section, we present a novel MIP model formulation of the profit-maximizing MDVRPTW. We introduce the full model in Section 3.1. In Sections 3.2 and 3.3, we then present two model variants which focus on maximizing the number of served customers and minimizing the routing costs, respectively. These variants allow us to isolate each of the previously discussed benefits of a flexible depot assignment. In Section 3.4, we provide further details of our implementation and formulate some valid inequalities that improve the computational performance of our models. We use the models formulated in this section in our numerical study in Section 4.

#### 3.1 The Profit-Maximizing MDVRPTW

In this section, we formally define the addressed planning problem and introduce and explain our corresponding mathematical formulation. Different from other exact formulations of the MDVRPTW in the literature (e.g., Engeler (2002), Bae and Moon (2016), Dondo et al. (2003), and Dondo and Cerdá (2007)), our formulation is a two-index formulation in that it does not involve indexes for vehicles or tours. This leads to a significantly smaller number of variables,
Sets
\(D\) set of depots
\(N\) set of customer locations

Parameters
\(Q\) homogeneous vehicle capacity
\(v_d\) number of vehicles at depot \(d\)
\([e_i, l_i]\) time window of customer \(i\)
\(t^0_{di}\) travel time between depot \(d\) and customer \(i\)
\(t_{ij}\) travel time between customers \(i\) and \(j\)
\(s_i\) service time of customer \(i\)
\(\delta_i\) demand of customer \(i\)
\(R\) maximal route duration
\(p_i\) revenue of customer \(i\)
\(M^A\) BigM for assignment constraints
\(M_i^Q\) BigM for capacity constraints
\(M_i^R\) BigM for route duration constraints
\(M_{ij}^T\) BigM for time window constraints

Variables
\(x_{ij}\) binary, 1 if customer \(j\) is visited directly after customer \(i\)
\(y_{di}\) binary, 1 if customer \(i\) is the first on a route from depot \(d\)
\(z_{di}\) binary, 1 if customer \(i\) is the last on a route from depot \(d\)
\(a_{ij}\) equal to the integer \(d\) of the depot customers \(i\) and \(j\) are assigned to if visited directly after each other, 0 otherwise
\(q_i\) cumulative demand served on a route after visiting customer \(i\)
\(\tau_i\) arrival time at customer \(i\)
\(r_i\) starting time at the depot of the route including customer \(i\)

Table 1: Notation used in the MIP model formulation

as well as less symmetry in the model, and thereby also to faster solution times. For example, Dondo and Cerdá (2007) state that their exact approach is able to solve an instance with 25 customers and two depots in about 24 minutes, whereas we can solve, e.g., the benchmark instances 1a (and 7a) of Cordeau et al. (2001) with four (six) depots and 48 (72) customers in five seconds (37 minutes).

Table 1 summarizes our notation. We consider an e-grocer that is using multiple depots \(d \in D\) to offer AHD. Given a set \(N\) of customer orders, she has to decide which orders to accept and how to design delivery routes to serve them such that the resulting profit is maximized. The profit consists of the revenues generated by accepting and delivering customer orders minus the costs for executing the delivery tours.

Within this setting, we make the following assumptions:

- For each customer \(i \in N\), the location, demand \(\delta_i\), revenue \(p_i\), time window \([e_i, l_i]\), and service time \(s_i\) are deterministic and given.

- All vehicles have an identical capacity of \(Q\).
• Each depot $d$ has a fixed number of vehicles $v_d$ available that start and end their tours at that depot.

• There is maximum $R$ on the allowable duration of a delivery tour.

• Travel costs are proportional to travel time. $t_{ij}$ denotes the travel time between customers $i$ and $j$, and $t_{0d}^d$ between depot $d$ and customer $i$.

We use the following decision variables in our model formulation. $x_{ij} \in \{0,1\}$ is equal to 1 if customer $j$ is visited directly after customer $i$ by a vehicle. Variables $y_{di} \in \{0,1\}$ and $z_{di} \in \{0,1\}$ indicate if a customer is the first or the last customer on a tour departing or ending at depot $d$, respectively. $q_i$ is the accumulated demand volume that is served between the depot and customer $i$ on the corresponding route. $r_i$ is the starting time at the depot of the tour visiting customer $i$. $\tau_i$ denotes the arrival time at customer $i$. $a_{ij}$ reflects the customer-depot assignment of customers $i$ and $j$. If customer $j$ is visited directly after customer $i$ on the same tour, $a_{ij}$ equals the integer value $d$ of the depot the two customers are assigned to, otherwise $a_{ij}$ equals 0. This special formulation enables us to keep track of the depot assignment without having to introduce a binary variable.

\[
\begin{align*}
\text{Max} & \quad \sum_{i \in \mathcal{N}} p_i \left( \sum_{d \in \mathcal{D}} y_{di} + \sum_{j \in \mathcal{N}} x_{ji} \right) \\
& \quad - \sum_{i \in \mathcal{N}} \left( \sum_{d \in \mathcal{D}} t_{0d}^d (y_{di} + z_{di}) + \sum_{j \in \mathcal{N}} t_{ij} x_{ij} \right) \\
\text{s.t.} & \quad \sum_{d \in \mathcal{D}} y_{di} + \sum_{j \in \mathcal{N}} x_{ji} \leq 1 \quad \forall i \in \mathcal{N} \quad (2) \\
& \quad \sum_{d \in \mathcal{D}} z_{di} + \sum_{j \in \mathcal{N}} x_{ij} = \sum_{d \in \mathcal{D}} y_{di} + \sum_{j \in \mathcal{N}} x_{ji} \quad \forall i \in \mathcal{N} \quad (3) \\
& \quad \sum_{i \in \mathcal{N}} y_{di} \leq v_d \quad \forall d \in \mathcal{D} \quad (4) \\
& \quad \sum_{i \in \mathcal{N}} z_{di} = \sum_{i \in \mathcal{N}} y_{di} \quad \forall d \in \mathcal{D} \quad (5) \\
& \quad \tau_i + s_i + t_{ij} - M^T_{ij} (1 - x_{ij}) \leq \tau_j \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \quad (6) \\
& \quad \sum_{d \in \mathcal{D}} t_{0d}^d y_{di} \leq \tau_i \quad \forall i \in \mathcal{N} \quad (7) \\
& \quad \tau_i \geq c_i \quad \forall i \in \mathcal{N} \quad (8) \\
& \quad \tau_i \leq l_i \quad \forall i \in \mathcal{N} \quad (9) \\
& \quad q_i \geq \delta_i \quad \forall i \in \mathcal{N} \quad (10) \\
& \quad q_i \leq Q \quad \forall i \in \mathcal{N} \quad (11) \\
& \quad q_i + \delta_j - M^Q_{ij} (1 - x_{ij}) \leq q_j \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \quad (12) \\
& \quad \sum_{d \in \mathcal{D}} d y_{di} + \sum_{j \in \mathcal{N}} a_{ji} = \sum_{j \in \mathcal{N}} a_{ij} + \sum_{d \in \mathcal{D}} d z_{di} \quad \forall i \in \mathcal{N} \quad (13)
\end{align*}
\]
\[ a_{ij} \leq M^A x_{ij} \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \] (14)

\[ r_j \leq r_i + M^R_R (1 - x_{ij}) \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \] (15)

\[ r_i \leq \tau_i - \sum_{d \in \mathcal{D}} t_{di}^0 y_{di} \quad \forall i \in \mathcal{N} \] (16)

\[ \tau_i + \sum_{d \in \mathcal{D}} t_{di}^0 z_{di} + s_i - r_i \leq R \quad \forall i \in \mathcal{N} \] (17)

\[ x_{ii} = 0 \quad \forall i \in \mathcal{N} \] (18)

\[ x_{ij} \in \{0, 1\} \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \] (19)

\[ y_{di} \in \{0, 1\} \quad \forall d \in \mathcal{D}, i \in \mathcal{N} \] (20)

\[ z_{di} \in \{0, 1\} \quad \forall d \in \mathcal{D}, i \in \mathcal{N} \] (21)

\[ a_{ij} \geq 0 \quad \forall i \in \mathcal{N}, j \in \mathcal{N} \] (22)

\[ r_i \geq 0 \quad \forall i \in \mathcal{N} \] (23)

Equation (1) maximizes the profit consisting of the revenue generated by delivering the customer orders and the costs for the delivery tours. Constraints (2) ensure that each customer is visited at most once. Constraints (3) make sure that the same number of vehicles leaves a customer location as arrives there. Constraints (4) and (5) ensure that at most \( v_d \) vehicles leave depot \( d \) and that the same number of vehicles returns after finishing the tours. Constraints (6) – (9) make sure that the time windows of all customers are respected. Constraints (10) – (12) ensure that the vehicle capacity is not exceeded.

Since we do not use an index for the vehicles or tours, we need a different way of ensuring that each vehicle returns to the depot it started from. We do so by using the customer-depot assignment expressed by \( a_{ij} \). We ensure that customers that belong to the same tour are assigned to the same depot. If customer \( i \) is the first on a tour, the first sum on the left-hand side of constraint (13) takes the value of the depot from where the tour has started, while the second sum is equal to zero. The right-hand side of the constraint works equivalently for the last customer on a tour. For intermediate customers, the first sum on the left-hand side and the last sum on the right-hand side are equal to zero. Then both sides indicate the customer-depot assignment of the direct predecessor or successor on the tour. Constraint (14) sets all \( a_{ij} \) to zero if \( j \) is not the direct successor of \( i \).

Constraints (15) – (17) ensure that the maximum route duration is respected. Again, since we do not have an index for the vehicles, we need to account for the tour length for each customer. We do this by denoting the starting time of the tour the customer belongs to by \( r_i \) and pass this value on in the tours through constraints (15). The starting time is set by (16) if the customer is the first on the tour. (17) then calculates the duration by adding the times needed to finish the tour to the arrival time at the node and subtracting the starting time. This must be smaller than the given maximum duration.

Finally, (18) forbids single-customer cycles, and (22) and (23) define the domains of the variables.
3.2 Maximizing the Number of Served Customers

If the ordering phase is considered in isolation, the objective is to maximize the number of served customers while constructing feasible delivery tours (Hungerländer et al., 2018). We can easily adapt the model (1) – (23) to this setting by substituting objective function (1) with:

$$\text{Max} \quad \sum_{i \in \mathcal{N}} \left( \sum_{d \in \mathcal{D}} y_{di} + \sum_{j \in \mathcal{N}} x_{ji} \right)$$ (24)

We use this model variant to assess the pooling benefit in Section 4.3.

3.3 Minimizing Routing Cost

When the route planning in the preparation phase is considered in isolation, the set of customers can be regarded as given from the ordering phase. The focus then is on minimizing the routing costs (Hungerländer et al., 2018). We can transform the model (1) – (23) in order to minimize the routing cost. The model is then equivalent to those presented in the literature having a pure routing perspective on the MDVRPTW, as e.g., Cordeau et al. (2001). For this, we have to replace objective function (1) and sets of constraints (2) and (3) with

$$\text{Min} \quad \sum_{i \in \mathcal{N}} \left( \sum_{d \in \mathcal{D}} t_{0d} (y_{di} + z_{di}) + \sum_{j \in \mathcal{N}} t_{ij} x_{ij} \right)$$ (25)

$$\sum_{d \in \mathcal{D}} y_{di} + \sum_{j \in \mathcal{N}} x_{ji} = 1 \quad \forall i \in \mathcal{N}$$ (26)

$$\sum_{d \in \mathcal{D}} z_{di} + \sum_{j \in \mathcal{N}} x_{ij} = 1 \quad \forall i \in \mathcal{N}$$ (27)

We make use of this model variant for analyzing the routing benefit in Section 4.2.

3.4 Model Implementation

In order to improve the computational performance of the above model, we use the following definitions of the BigMs in our numerical experiments and introduce additional valid inequalities.

To reduce the impact of the BigM constraints, we use constraint-specific large numbers. In particular, we set:

$$M_A^A = \max_{d \in \mathcal{D}} d \quad \text{assuming all } d \geq 0$$ (28)

$$M_i^Q = Q + \delta_i \quad \forall i \in \mathcal{N}$$ (29)

$$M_i^R = \max\{l_i - \min_{d \in \mathcal{D}} t_{0d}, 0\} \quad \forall i \in \mathcal{N}$$ (30)

$$M_{ij}^T = l_i + s_i + t_{ij} \quad \forall i \in \mathcal{N}, j \in \mathcal{N}$$ (31)
Additionally, we introduce the following valid inequalities that help to speed up the computations:

\[
\sum_{i \in N} \delta_i \left( \sum_{d \in D} y_{di} + \sum_{j \in N} x_{ji} \right) \leq Q \sum_{d \in D} \sum_{i \in N} y_{di} \tag{32}
\]

\[
\sum_{i \in N} \left( \sum_{d \in D} t_{0_{di}} (y_{di} + z_{di}) + \sum_{j \in N} t_{ij} x_{ij} + s_i \left( \sum_{d \in D} y_{di} + \sum_{j \in N} x_{ji} \right) \right) \leq R \sum_{d \in D} \sum_{i \in N} y_{di} \tag{33}
\]

These constraints can be interpreted as upper bounds on the number of customers that can be served, from a vehicle capacity perspective (32) and from a route duration perspective (33).

For the model maximizing the number of served customer, the BigM definitions as well as the valid inequalities can be applied without changes. However, for the model that minimizes the routing costs, we rewrite (32) as

\[
\sum_{d \in D} \sum_{i \in N} y_{di} \geq \left\lceil \sum_{i \in N} \delta_i \frac{Q}{Q} \right\rceil \tag{34}
\]

This way we get a tighter lower bound on the number of tours needed from a vehicle capacity perspective. Then (33) can be interpreted as a lower bound on the number of tours required.

For all model versions, it is easy to cut down the model size by excluding all \( x_{ij} \) and corresponding constraints for which \( e_i + s_i + t_{ij} > l_j \). The indices of the sums in the constraints then have to be modified accordingly.

### 4 Analysis of the Benefits of Flexibility

In this section, we assess the potential benefits of a flexible depot assignment in a numerical study. We start by separately investigating each of the benefits described in Section 2. Specifically, we analyze the routing benefit in Section 4.2 and the pooling benefit in Section 4.3. Subsequently, we consider the interaction between both effects and investigate the overall profit benefit in Section 4.4. Before presenting our results, we explain the setup of our numerical experiments and the solution procedure in the following subsection.

#### 4.1 Experimental Setup

Throughout our experiments, we consider a square service region of size 100 \( \times \) 100 (see Figure 2). Within this setup, we analyze multiple scenarios, characterized by different parameter settings and generate multiple random instances for each scenario. Table 2 lists the parameters and the values considered. We start by analyzing a base case scenario, highlighted in boldface.

For each scenario, we consider the case of two depots, located at the centers of the upper and lower half of the service region (see Figure 2) and the case of four depots, located at the centers of the four quadrants of the service region. We consider a fixed fleet of vehicles assigned
Table 2: Parameter settings considered in the experiments; **bold face** indicates the base case settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values considered</th>
</tr>
</thead>
<tbody>
<tr>
<td># Depots</td>
<td>2; 4</td>
</tr>
<tr>
<td>Customer layout</td>
<td><strong>uniform</strong>: clustered towards / away from the depot</td>
</tr>
<tr>
<td>Demand / capacity ratio</td>
<td>1: 0.90, 0.93,..., 1.20</td>
</tr>
<tr>
<td>Shift length</td>
<td><strong>500</strong>: 100</td>
</tr>
<tr>
<td>Time window length</td>
<td><strong>0.25 × shift length</strong>: 0.125 × shift length; none</td>
</tr>
<tr>
<td>Time window clustering</td>
<td><strong>random</strong>: clustered</td>
</tr>
</tbody>
</table>

to each depot. In the base case, we assume four vehicles with a capacity of eight that are equally split over the depots. We generate customer requests based on different customer layouts. To reflect different time-window offerings, we subdivide each quadrant of the service region into four square sectors (labeled A-D in Figure 2). Unless stated otherwise, we assume the same number of customers in each sector and draw their coordinates uniformly distributed within the sector. We repeat this step to generate 25 random instances. In subsequent experiments, we consider clustered customer layouts. Order sizes are homogeneous of unity size. In the base case, we assume that total demand equals the total capacity of the vehicle fleet. In the analysis of the pooling benefit, we vary the demand-capacity ratio between 0.9 and 1.2.

Vehicles travel at a constant speed of one distance unit per time unit. To control the impact of time scarcity, we introduce a shift length parameter that reflects the maximum allowable route duration. We assume a shift length of 500 in the base case and lower this value to 100 in more time-constrained scenarios. In addition, the shift length represents the time interval covered by the offered time windows. We consider the cases of four (base case) and eight time windows of equal length. We also consider different patterns of time window assignments to customers. In the base case, we draw a random time window for each customer independently. This corresponds to a setting in which customers have a wide choice of available time windows.

![Illustrative test instance of the base case with 32 customers and two depots. Letters A–D indicate the sectors for the time window clustering used in Section 4.2.3](image-url)
Alternatively, we consider clustered time-window assignments to the 16 sectors of the service area (see Figure 2) to mimic the case of e-grocers that steer demand through more restrictive time window offerings.

In our experiments, we solve the generated problem instances by using the MIP formulations introduced in Section 3. To this end, we first use the respective model variant to calculate a decomposed solution in which we fix the assignment of each customer to its closest depot and solve the problem for each depot separately. We then use this solution as a warm start for the MDVRPTW with a flexible depot assignment. When analyzing the routing costs in Section 4.2, we use the VNS heuristic of Polacek et al. (2004) in order to find a good starting solution for the MIP model of the decomposed problem as well as to improve the starting solution for the MDVRPTW. In order to run a large number of experiments, we limit the computation times and VNS iterations for each problem instance as stated in Table 3. We implemented our solution approaches in Python and used Gurobi 8.0.0 as a MIP solver on a Lenovo notebook with an i5-5300 CPU at 2.3 GHz, 8 GB RAM and 64 bit Windows 8.1.

### 4.2 Routing Benefit

As discussed in Section 2, a flexible customer-to-depot assignment can potentially reduce routing costs. In this section, we aim to quantify this benefit for different contexts, as defined by the customer and depot geography and the service time windows. To isolate the routing benefits, we use the model variant of Section 3.3. We thus consider the routing decisions for a given set of accepted customers.

Table 4 summarizes our base case results. We observe that a flexible depot assignment is associated with substantial savings in routing costs as compared to the case in which each customer is always served from the nearest depot. Calculated savings amount to up to 13.4%. We further observe that savings are higher for four depots than for two depots, with average savings of 6.4% and 4.4%, respectively.

To better understand the mechanism behind the savings, we compare the actual routing solutions in the different scenarios. Figure 3 visualizes the differences between the solutions. The figure shows the customer locations of all 25 base case instances and highlights which

<table>
<thead>
<tr>
<th>Routing benefit</th>
<th>2 depots</th>
<th>4 depots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>4.4%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Maximum</td>
<td>8.0%</td>
<td>13.4%</td>
</tr>
</tbody>
</table>

Table 4: Routing benefit in the base case instances
Figure 3: The pictures show the generated customer locations for all base case instances, with two depots on the left and four depots on the right. Black dots indicate customers that are served from a depot other than the closest one in at least one instance under a flexible depot assignment. The empty circles indicate all other customers. The shading of the area reflects the value of the introduced relative distance measure.

As to be expected, the customers that switch between depots are mainly located around the boundaries of the delivery regions of different depots. To formalize this observation, we show different shades in the graph that correspond to different relative distances, where the relative distance for customer \( i \) is defined as, \( \text{relDist}_i = \frac{\text{dist}_{1i}}{\text{dist}_{1i} + \text{dist}_{2i}} \), with \( \text{dist}_{1i} \) being the distance the closest depot and \( \text{dist}_{2i} \) the distance to the second closest depot. We see that most switching between depots occurs in the area with a high relative distance.

4.2.1 Effect of the Customer Layout

Intuitively, the customer layout affects the achievable routing benefits. To assess the impact of this factor, we consider instances in which customer locations are not uniformly distributed. Specifically, we repeat our experiments on two sets of instances with clustered customer locations. In the first case, customers are clustered around the centers of the four quadrants of the service region (‘towards the depots’). In the second case, they are clustered towards the center of the service region (‘away from the depots’). In both cases, we draw the coordinates of the customer locations from a beta distribution to achieve the clustering. We again generate 25 instances for each of the clustering variants.

Figure 4 summarizes the obtained routing benefits for the different ways of clustering. We
observe that the routing benefits decrease (increase) for a clustering of customers towards (away from) the depots. We can use our previously defined relative distance metric to explain this effect, as illustrated in Figure 5. When customer locations are clustered towards the depots there are few customers in the areas with a high relative distance. With less customer on the boundary region between two depots, there are less opportunities to exploit the additional flexibility in the depot assignments. The opposite is true for a clustering of the customers away from the depots.

### 4.2.2 Effect of the Depot Location

Complementary to modifying the customer distribution, we also consider the impact of changing the depot location. This is relevant since depots cannot always be installed in the center of a delivery region, e.g. due to high real estate costs or unavailability of appropriate sites. To investigate the impact of such constraints, we resolved the base case instances with two depots for modified depot locations. We leave the vertical coordinates of the depots unchanged but set the horizontal coordinate first to 25 and then to 0. Figure 6 illustrates the results. Moving the depots out of the center of the service region increases the area with a high value of the relevant distant measure. Correspondingly, we observe a large number of customer reassignments in the flexible solution. As a result, the average routing benefits increase to 4.6% and 5.3% for a horizontal depot coordinate of 25 and 0, respectively. The corresponding maximum values increase to 9.2% and 12.7%, respectively.
Figure 5: These pictures illustrate, for different customer layouts, where customers are located and which customers (black dots) are served from a depot other than the closest one in at least one instance under a flexible depot assignment.
4.2.3 Effect of Time Windows

Time windows are another factor that one would expect to affect the routing benefit of a flexible depot assignment. In this section, we explore the impact of this factor. To this end, we modify the time window settings in our base case instances. We jointly vary two factors, namely the length of the time windows and their assignment to the customers. First, we consider time windows that are half as long as in the base case. Second, we replace the random assignment of time windows to customers by a clustered time window assignment, where all customers in a sector (illustrated with A–D in Figure 2) receive the same long time window or adjacent short time windows. The latter reflects a setting where an e-grocer is able to steer customers to favorable time windows during the ordering phase. Finally, we also consider the case of no time windows at all.

Figure 7 illustrates the resulting routing benefits for the different time window settings, again for two and for four depots. As for the base case, the benefits are systematically higher for four depots than for two depots. Regarding the time window setting, the biggest impact comes from the time window clustering. Across all settings, a geographical clustering of time windows reduces the average routing benefit by roughly 50%, relative to a random time window assignment. A potential explanation is that in the clustered setting, a customer’s time window may be incompatible with the schedule of a vehicle from a different depot. In comparison, the length of the time windows has a smaller impact on the routing benefits in our results. While the average benefit remains almost unaffected, the maximum benefit increases for shorter time windows. Finally, the effects of removing time windows entirely are mixed. For two depots, the average routing benefits under no time windows are similar to those under clustered
time windows whereas for four depots they are similar to those under randomly assigned time windows.

In conclusion, our results show that time-window management and a flexible depot assignment act as substitutes. If an e-grocer is able to steer demand by means of a tailored time-window offering the additional value of flexibility in the depot assignment is limited. On the other hand, a flexible depot assignment helps e-grocers to limit the operational cost implications of offering a broader choice of time-windows to customers.

### 4.3 Pooling Benefit

In the previous section, we considered the routing benefits for a given set of accepted customer orders. In this section, we focus on the benefits that arise in the customer acceptance phase by pooling the vehicle capacity of different depots. To assess the pooling benefit, we track the number of additional customers served in the flexible depot setting as compared to the fixed depot setting. As before, we use our MIP model to solve the different problem instances. However, instead of minimizing the routing costs, we now maximize the number of accepted customers as presented in Section 3.2.

In contrast to the previous section, capturing pooling effects requires us to consider instances with varying demand volumes in our experiments. We generate such instances in a two-step process. First, we randomly generate a uniformly distributed customer base of 96 customers. Second, in each instance, each of these customers places an order with probability $\pi$. In our experiments, we vary $\pi$ from 0.3 to 0.4. For a base system-wide capacity of 32, this corresponds to a demand-capacity ratio between 0.9 and 1.2. We generate 25 instances for each value of $\pi$. 

![Boxplots showing routing benefit for different time window variants](image)
Figure 8: For different values of the demand-capacity ratio (varying the order probability $\pi$), the figure shows the relative pooling benefit (the relative increase of customers served) between the flexible assignment and the fixed setting. For a shift length of 500, the figure compares the different time window settings obtained in our experiments with values of the pure pooling benchmark without routing effects.

As a benchmark, we consider the special case of no time windows and unconstrained route duration. In this case, the number of served customers is only limited by the total capacity of the fleet of vehicles. This means that the benefits in this setting can be calculated analytically, without solving a routing problem. Specifically, we can express the expected number of served customers $V$ for the flexible depot setting as: $V(n, Q^T) = \sum_{k=1}^{n} \min\{k, Q^T\} \binom{n}{k} \pi^k (1-\pi)^{n-k}$ for a customer base of $n$ customers and a system capacity of $Q^T$. In the fixed case with $r$ depots, the number of served customers then equals $rV(\frac{n}{r}, \frac{Q^T}{r})$.

Against this benchmark, we consider various time-constrained settings that do require an instance-specific routing analysis. To this end, we assume a shift length of $R = 500$, in combination with the different time-window settings introduced in Section 4.2.3.

Figure 8 shows the relative increase in the expected number of customers served due to capacity pooling, both for the benchmark and for the different time-constrained settings, for two and four depots. For the pure pooling benchmark, we observe that most benefits can be achieved when expected demand is equal to the total system capacity. The reason is that depot flexibility is less useful if all customers can be served from their closest depot (low demand-capacity ratio) or if the system-wide capacity is exhausted (high demand-capacity ratio). Moreover, we observe that the benefits increase with the number of depots. That is, we see benefits of up to 2.5% with two depots and up to 6% with four depots. This reflects the famous ‘square root law’ in inventory theory (cf. e.g. Chopra and Meindl, 2013, p. 343). The results of the time-constrained instances follow these patterns of the benchmark curve rather closely. In general, the benefits are slightly lower than for the pure pooling benchmark. This suggest that the interactions between
the different constraints generally limit the benefits of flexible-depot strategy. That is, there may be less beneficial switching options between depots due to incompatible time windows. However, we also observe that, in a few cases, benefits are even larger in the time constrained instances. This is because the temporal constraints add another capacity dimension that can be ‘pooled’ over the different depots. That is, it can happen that within a certain time window all vehicles are used in one depot but still available in another.

Figure 9 show the same results for a setting in which we further tightened the time constraints by reducing the shift length to $R = 100$. Here, we see that the clear pattern of the physical capacity disappears as the time constraints become the dominating factor. The reason is that the time constraints affect both the solutions of the fixed and of the flexible depot assignment, and it depends on the specific instance which solution is affected more severely. As a result, the relative pooling benefit in the 2-depot case may even exceed the one in the 4-depot case in some instances.

4.4 Combined Benefit: Profit

Having gained insights into both the routing cost benefit and the capacity pooling benefit of a flexible depot assignment we now address their interaction by returning to the general problem described in Section 2. That is, we combine the routing and order acceptance decisions in the MIP model defined in Section 3.1 and maximize total profit, which includes revenues for the customers served and routing costs for the delivery tours.

We assume a constant revenue $p$ per customer and calibrate our scenarios by setting the revenues roughly equal to the routing costs per customer obtained in Section 4.2, i.e. $p = 20$. In our experiments, we then vary $p$ between 15 and 30. For generating the instances, we use the same logic as in Section 4.3. The results are summarized in Table 5.
<table>
<thead>
<tr>
<th>Revenue per customer</th>
<th>Depots</th>
<th>Shift Length</th>
<th>TW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆ profit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆ cust</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆ cpc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>2.49</td>
<td>1.93</td>
<td>2.21</td>
</tr>
<tr>
<td>lc</td>
<td>0.50</td>
<td>2.84</td>
<td>1.86</td>
</tr>
<tr>
<td>lr</td>
<td>1.11</td>
<td>0.89</td>
<td>1.12</td>
</tr>
<tr>
<td>sc</td>
<td>0.00</td>
<td>3.58</td>
<td>2.03</td>
</tr>
<tr>
<td>sr</td>
<td>0.00</td>
<td>0.48</td>
<td>1.94</td>
</tr>
<tr>
<td>100</td>
<td>2.96</td>
<td>2.09</td>
<td>1.36</td>
</tr>
<tr>
<td>lc</td>
<td>2.46</td>
<td>1.68</td>
<td>1.90</td>
</tr>
<tr>
<td>lr</td>
<td>2.18</td>
<td>2.43</td>
<td>3.00</td>
</tr>
<tr>
<td>sc</td>
<td>2.71</td>
<td>1.67</td>
<td>1.95</td>
</tr>
<tr>
<td>sr</td>
<td>0.07</td>
<td>2.45</td>
<td>1.53</td>
</tr>
<tr>
<td>500</td>
<td>6.08</td>
<td>5.89</td>
<td>3.99</td>
</tr>
<tr>
<td>lc</td>
<td>1.08</td>
<td>4.74</td>
<td>5.24</td>
</tr>
<tr>
<td>lr</td>
<td>0.08</td>
<td>2.94</td>
<td>3.76</td>
</tr>
<tr>
<td>sc</td>
<td>0.41</td>
<td>3.28</td>
<td>4.09</td>
</tr>
<tr>
<td>sr</td>
<td>0.00</td>
<td>1.09</td>
<td>1.83</td>
</tr>
<tr>
<td>20</td>
<td>10.32</td>
<td>6.61</td>
<td>5.72</td>
</tr>
<tr>
<td>lc</td>
<td>1.94</td>
<td>5.64</td>
<td>5.38</td>
</tr>
<tr>
<td>lr</td>
<td>1.67</td>
<td>7.66</td>
<td>9.18</td>
</tr>
<tr>
<td>sc</td>
<td>1.30</td>
<td>5.53</td>
<td>5.32</td>
</tr>
<tr>
<td>sr</td>
<td>1.06</td>
<td>8.79</td>
<td>9.14</td>
</tr>
</tbody>
</table>

Table 5: Shows the relative change (in %) from a fixed to a flexible assignment of the values profit (Δ profit), number of served customers (Δ cust), and routing cost per served customer (Δ cpc), for different values of the revenue per customer and for the time window variants: no, long clustered (lc), long random (lr), short clustered (sc), and short random (sr).
We quantify the benefits as the relative increase in total profit of the flexible customer-to-depot assignment relative to the fixed depot assignment. To reflect the components of the profit function, we also consider the relative change in the number of customers served and in the cost per customer. Table 5 shows the relative benefits for different values for the unit revenues, different numbers of depots, different shift lengths and different time window variants. The resulting overall picture is complex and depends on the intricate interaction of the various design parameters. However, we can observe a few systematic patterns.

First, as before, we see that the relative benefits tend to be larger for the instances with four depots than for those with two depots. We also see that the benefits tend to increase with the shift length.

Furthermore, we see that we systematically accept more customers in the flexible-depot case than in the fixed-depot case. These additional customers imply higher revenues. The effects on the cost side are more diverse. From Table 5, we see that the cost per customer served increases in many cases, but does not outweigh the revenue increase. However, routing costs per customer may actually decrease through a flexible depot assignment (see bold entries in the table). We mainly observe this effect for high unit revenues. In this setting, maximizing the number of customers served becomes the dominant objective. The flexible depot assignment may be able to achieve this objective with more efficient routes than the fixed assignment. That is, we get routing benefits similar to those in Section 3.3. In general however, the revenue increase is the dominant component of the profit increase.

In Table 5, we further observe that the relative benefits tend to decrease with the unit revenues. Note, however, that this involves a scale effect. As unit revenues increase so do total profits, ceteris paribus. In addition, it becomes optimal to accept more customers, both in the fixed and in the flexible depot-assignment case. Moreover, a larger number of profitably reachable customers increases the optimization potential of the flexible assignment. Thus the absolute benefit of flexibility also tends to increase with the unit revenues, but in many cases less fast than total profits. This results in decreasing relative benefits even though absolute benefits are growing.

The interaction between the different time window settings, shift lengths and other capacity considerations results in additional more subtle effects. What we observe, in general, is that the value of flexibility first increases when the setting becomes more constrained but then decreases when it becomes too constrained. On one end of the spectrum, without strict time or capacity constraints, there is not much to be gained from a flexible depot assignment as the fixed assignment already performs very well. On the other end of the spectrum, the problem is so constrained that there is no room for the flexible-assignment strategy to improve over the fixed-depot strategy.

To illustrate this point, Figure 10 displays the average profits of both depot assignment approaches for different time window settings, for a particular parameter combination (four depots, shift length = 500, \( p = 30 \)). We see that average profits decrease as the time window settings are constraining the problem more severely (moving from no time windows to clustered and to random ones and from long to short time windows). However, the profit gap between
the depot assignment strategies is non-monotone but reaches a maximum for long random time windows.

We observed similar patterns for other constraining factors, e.g., for the shift length. Where exactly the region of the largest benefits is located depends on the specific parameter settings. This insight is important since it means that assessing the potential of a flexible depot assignment requires a good understanding of the effective constraints in that situation.

5 Conclusion

Motivated by observations in the e-groceries sector, this paper considers a setting in which a service provider makes deliveries to customers subject to time windows from a number of depots. In this setting, we study the advantages of flexibly assigning customers to depots as compared to always serving all customers from their nearest depot. We identify two types of benefits associated with a flexible-depot assignment: (i) the routing benefits of centrally constructing the delivery routes for the entire delivery region and (ii) the pooling benefit of pooling the vehicle fleet capacities across different depots. The routing benefits result in costs savings while the pooling benefits increase the number of served customers, given the available capacity at each depot.

While a flexible depot assignment has several potential benefits, it is more complex to plan and execute than a simple fixed plan. Therefore, it is relevant to investigate under what conditions the flexible assignment provides most benefits. To that end, we conducted an extensive numerical study. In our experiments, we found average profit increases in the range of 3% to 5%.
For many online grocery businesses which are notoriously struggling to attain profitability this is a very relevant order of magnitude. In addition, we found a large variability of the benefits across scenarios, with maximum benefits exceeding 10%.

Naturally, the benefits of a flexible depot assignment depend both on the network of depots and on the customer layout. Overall benefits increase in the number of depots. Moreover, the routing benefits are larger when there are more customers close to the boundaries of the different delivery regions. We also found larger routing benefits when time windows are randomly distributed across customers than when they are geographically clustered. This suggests that flexibility in the depot assignment may help an e-grocer to offer a broader menu of time windows to its customers.

The pooling benefits are largest if the expected demand is roughly equal to the available capacity. Again, this is a practically relevant setting, since e-grocers strive for high capacity utilization.

In general, we found that the main benefit of a flexible depot assignment lies in the ability to profitably serve more customers. Again, this is in line with the growth objectives of many e-grocers. In response, the cost per customer may increase or decrease, depending on the specific setting.

Finally, our results clearly document that flexibility in the depot assignment is not a panacea. In a very loosely constrained setting, this flexibility has little value. Yet, such a loosely constrained setting is not what we typically observe in reality. On the other hand, flexibility in the depot assignment also adds little value in very highly constrained settings. However, for intermediate settings, i.e. for companies operating ‘on the edge’, a flexible depot assignment may provide just the additional leeway required to substantially enhance profits.

Our results in this paper provide insights into the magnitude of the benefits that can be derived from a flexible depot assignment in home-delivery e-fulfillment operations and into the factors that drive these benefits. In future research, it will be interesting to extend our analysis, e.g., to other geographical setups of the service region or to other time-window menus. In particular, it would be insightful to complement our simulation results with analyses of specific business cases. Furthermore, our analysis highlights the close interconnection between the ordering phase and the preparation phase in e-fulfillment. Research on this interaction is still scant and offers much potential for relevant future contributions.
References


